

# **Kollokationen als zentrales Übersetzungsproblem — Vorschläge für eine Kollokationsdatenbank Deutsch-Französisch/Französisch-Deutsch auf der Basis paralleler und vergleichbarer Korpora**

Kathrin STEYER

*Institut für deutsche Sprache Mannheim*

## **1 Das idiomatische Prinzip**

Übersetzen ist nur zu einem geringen Prozentsatz eine Wort-für-Wort-Transformation. Die Experten sind sich heute einig, daß beim Übersetzen und Formulieren fremdsprachiger Texte in der Regel — mit Ausnahmen von strikt terminologisch bestimmten Übersetzungen — größere Übersetzungseinheiten abgerufen und verwendet werden: Mehrwortkonstruktionen, Kollokationen, idiomatische Wendungen usw. Hausmann bezeichnet diese sprachlichen Gebilde als "Halbfertigprodukte der Sprache" (vgl. Hausmann 1985: 118). Im Vorwort zu Langenscheidts Kontextwörterbuch Französisch-Deutsch beschreibt er die Problematik dieser sprachlichen Bausteine folgendermaßen:

Die Kombinierbarkeit der Wörter, ihre Fähigkeit, miteinander Verbindungen (oder Fügungen) einzugehen, ist demnach begrenzter als es uns bewußt ist. [...] Beim muttersprachlichen Wortgebrauch ist es, als gingen wir in schwindelnder Höhe nachtwandlerisch über einen schmalen Grad, ohne Bewußtsein der Abgründe links und rechts. Der Weg erscheint uns leicht und ungefährlich; alles ist selbstverständlich. Ganz anders der Fremdsprachler, der unweigerlich und ständig in Absturzgefahr ist. (Hausmann 1993: 5)

Hinter dieser metaphorischen Beschreibung verbirgt sich eines der schwierigsten Probleme, die sich bei der Produktion von Äußerungen und Texten in einer fremden Sprache ergeben. Es reicht nicht aus, die lexikalischen und grammatischen Regeln einer Sprache zu beherrschen. Vielmehr ist die 'idiomatische Kompetenz' des Sprechers ausschlaggebend für die kulturelle Angemessenheit seiner fremdsprachigen Äußerungen. Eine direkte Kausalbeziehung zwischen der idiomatischen Kompetenz und wirklichem interkulturellem Verstehen stellt Stubbs her (vgl. Stubbs 1997). Dabei ist für ihn das idiomatische Prinzip in der Sprache viel weitreichender als das bisher in der traditionellen Phraseologie- und Idiomatikforschung be-

schrieben wurde. Er verweist auf die Ansätze des britischen Kontextualismus der 30er Jahre, die bereits gezeigt haben, daß Wörter nicht frei gewählt werden könnten, daß sie in ihren Kombinationsmöglichkeiten eingeschränkt seien und in typischen Kombinationen verwendet würden (ebd. : 156). Ein Ergebnis der Korpuslinguistik der letzten Jahre sei die Erkenntnis, daß ein hoher Anteil des Sprachgebrauchs durch vorgefertigte Äußerungen stark routiniert ist. Dabei, so Stubbs, handele es sich kaum um Idiome im traditionellen Sinne. Die Sätze seien zwar feste Wortverbindungen, aber mit einer transparenten Bedeutung, also konventionell und idiomatisch, aber keine Idiome (ebd. : 153). Ihre Bestandteile kommen häufiger miteinander vor als andere. Sie stellen Standardverwendungen, typische Kombinationen — also zentrale Verwendungen — dar (z.B. häufig wiederkehrende Nomen-Verb- bzw. Nomen-Adjektiv-Verbindungen, standardisierte Verbalphrasen usw.).

Bei solchen Kombinationen geht es um Wahrscheinlichkeiten, Erwartungen und quantitative Verteilungen. Es geht um Normen des Sprachgebrauchs. Hier wird die Verbindung zwischen Idiomatik und kultureller Kompetenz deutlicher: man muß Versprachlichungen kulturell festgelegter Konzepte (z.B. Sprechakte), also landeskundlicher Informationen, beherrschen, um sich idiomatisch ausdrücken zu können. Durch den Gebrauch von idiomatischen Redewendungen werden gemeinsame Wertvorstellungen abgerufen. Nur wer die idiomatischen Ausdrücke und die damit verbundenen Wertungen oder Vorstellungen in einer Kultur kennt, kann also Texte verstehen, bzw. erfolgreich kommunizieren. (Stubbs 1997 : 157)

Will man zentrale Verwendungen von Wortverbindungen, also usuellen Gebrauch, erfassen, liegen die Konsequenzen auf der Hand (auch Stubbs zieht diese Konsequenz): Nur umfangreiche Korpora können Aufschluß über das geben, was wirklich gesprochen wird, was Usus ist. In bezug auf zwei- und mehrsprachige Kommunikation bedeutet dies, daß dem Übersetzer bzw. dem Produzenten eines fremdsprachigen Textes ein Werkzeug in die Hand gegeben werden muß, das es ihm ermöglicht, schnell und angemessen solche Wortverbindungen mit ihren typischen Struktureigenschaften und Verwendungsweisen abzurufen. Es müssen also systematische, geordnete und schnell abrufbare Listen dieser üblicherweise geltenden Wortverbindungen, ihrer typischen Verwendungsweisen und Kontexte erstellt werden. Einen Weg dahin eröffnet die Aufbereitung und Speicherung dessen, was bereits übersetzt wurde. Umfangreiche Datenbanken schon übersetzter Texte können in der Tat dazu dienen, jene standardisierten Wortverbindungen quasi als Bausteine aufzufinden und bei der Textproduktion einzusetzen.

Ein Kernkonzept dieser aufzubereitenden und abrufbar zu machenden Standardwortverbindungen stellen ohne Zweifel 'Kollokationen' dar, jene charakteristischen Wortkombinationen einer Sprache, die sich im Sinne von Hausmann im Spannungsfeld zwischen völlig freien Wortverbindungen einerseits und festen idiomatischen Redewendungen andererseits befinden :

La collocation se distingue de la combinaison libre (*the book is useful/das Buch ist nützlich/le livre est utile*) par la combinabilité restreinte (ou affinité) des mots combinés (*feuilleter un livre vs. acheter un livre*). La collocation se distingue

d'autre part des locutions (*idioms*, Redewendungen, par ex. *monter un bateau à qn/jdn. durch den Kakao ziehen/to pull sb's leg*) par son non-figement et par sa transparence. (Hausmann 1989: 1010)

Kollokationen können in besonderer Weise zu einem Übersetzungsproblem werden, da sie nicht in so hohem Maße standardisiert sind wie feste Redewendungen, aber auch nicht beliebig austauschbar. Beispielsweise kommt ein Substantiv gehäuft in Verbindung mit einem bestimmten Verb vor; ein anderes Verb wäre rein grammatikalisch nicht falsch und vielleicht auch nicht sinnentstellend, aber eben nicht üblich, weniger gebräuchlich. Da helfen dann die traditionellen Wörterbücher oft nicht weiter. Bevor dieses Problem der Kollokationen in zwei- und mehrsprachiger Kommunikation anhand einiger Beispiele diskutiert wird, soll kurz auf das Konzept 'Kollokation' eingegangen werden, das in der Linguistik sehr unterschiedlich in Abhängigkeit vom jeweiligen Forschungsinteresse verwendet wird.

## 2 Der Kollokationsbegriff — ein Integrationsvorschlag

Zahlreiche Arbeiten, die das Phänomen der 'Kollokation' in irgendeiner Weise thematisieren, betonen die Uneinheitlichkeit der Kollokationsauffassungen<sup>1</sup>.

Wir wollen uns an dieser Stelle nicht grundsätzlich mit diesen Konzepten auseinandersetzen, sondern vielmehr versuchen, verschiedene Herangehensweisen, die für uns keine unüberbrückbaren, sondern zusammengehörende Paradigmen einer komplexen Sprachbeschreibung sind, zu integrieren. Inhalte bzw. methodische Zugänge sollen dabei lediglich unter dem Aspekt betrachtet werden, ob sie in Hinblick auf mehrsprachige Nutzungsbedürfnisse praktikabel und verbindbar für eine übersetzungsbezogene Kollokationssystematik sind.

Viele Ansätze beschäftigen sich mit Kollokationen unter dem Aspekt der Regelhaftigkeit. Dabei geht es um die Frage, inwieweit die potentielle Kombinierbarkeit von Wörtern und damit verbundene Restriktionen aus dem Sprachsystem ableitbar sind bzw. inwieweit sie von der sprachlichen Norm bestimmt werden. Es werden also die norm- und regelgerechten Kombinationen von Wörtern einer Sprache untersucht, wobei sich die Analysen vor allem auf die syntaktische und semantische Kombinierbarkeit konzentrieren (auf syntagmatische bzw. paradigmatische Vereinbarkeitsbedingungen bzw. Selektionsbeschränkungen). Umfangreiche Forschungen gibt es auf dem

<sup>1</sup> Hier sei nur auf zwei aktuelle Arbeiten verwiesen (LEHR 1996, BAHNS 1996). Insbesondere Bahns gibt einen sehr instruktiven Überblick über die Geschichte der Kollokationsforschung von Firth bis Benson und Hausmann. Er ordnet auch Arbeiten anderer Forschungsrichtungen wie die 'lexikalischen Solidaritäten' (Coseriu) und die 'lexikalischen Funktionen' (Mel'cuk) in diesen Zusammenhang ein. Dabei verweist er besonders auf Benson, der ebenfalls eine Kollokationstypologie entwickelt hat und dem engen Kollokationsbegriff Hausmanns sehr nahe ist. Er präferiert diese Ansätze im Gegensatz zum seiner Meinung nach weiten Kollokationsbegriff des britischen Kontextualismus. (vgl. BAHNS 1996: 28-29)

Gebiet der Syntax und Valenz (vgl. hier u.a. Bresson/Kubczak 1998) und der semantischen Vereinbarkeitsregeln (vgl. dazu u.a. Viehweger 1977; 1982). Fast unberücksichtigt sind bisher die (ko-)textuellen Selektionsbeschränkungen, die aber doch einen wesentlichen Einfluß darauf haben, ob Wörter in einer bestimmten Art und Weise miteinander kombiniert werden können oder nicht.

Einen komplexeren Ansatz zur Definition von Kollokationen bietet Hausmann, der sie immer auch aus lexikographischer und fremdsprachen-didaktischer Perspektive betrachtet. Er versteht unter 'Kollokation' die typische, spezifische und charakteristische Zweierkombination von Wörtern, die er zur 'langue' zählt und die er von unspezifischen Wortverbindungen abgegrenzt haben will:

Eine Theorie der Kollokation wird vor allem zweierlei zu leisten haben. Sie muß einerseits die Kollokation als charakteristische Zweierkombination abgrenzen gegen unspezifische, banale Zweierkombinationen, die der parole und nicht der langue angehören. Zum zweiten muß sie den Status der beiden Kombinationspartner in dieser Zweierkombination zueinander untersuchen. (Hausmann 1985 : 118)<sup>2</sup>

Hinsichtlich des Status der Kombinationspartner unterscheidet er zwischen Basis und hinzutretendem Kollokator, die ein hierarchisches Verhältnis zueinander aufweisen, bei dem ein Partner determiniert (Basis) und einer determiniert wird (Kollokator). So nennt er das Beispiel *schütteres Haar*, bei dem *Haar* die Basis und *schütter* der Kollokator sei (ebd.:119)<sup>3</sup>. Hausmann vertritt dabei die Auffassung, daß das Nomen die wichtigste Basiswortart ist. Folgende Wortartkombinationen stellt er sich generell als Kollokationen vor :

On appellera collocation la combinaison caractéristique de deux mots dans une des structures suivantes :

- a) substantif + adjectif (épithète)
- b) substantif + verbe
- c) verbe + substantif (objet)
- d) verbe + adverbe
- e) adjectif + adverbe
- f) substantif + (prép.) + substantif. (Hausmann 1989 : 1010)

<sup>2</sup> Zur Illustration sei ein Beispiel aus einem Projekt des Instituts für deutsche Sprache in Mannheim (IDS) angeführt, bei dem eine Datenbank der wichtigsten Adjektiv-Substantiv-Verbindungen des Deutschen erstellt wurde. Durch diese Arbeit ist die Relevanz der Hausmannschen Unterscheidung zwischen spezifischen und unspezifischen Wortverbindungen sehr deutlich geworden. So ist die Adjektiv-Nomen-Verbindung *ein hübsches Sümmchen* eine typische (spezifische) Kombination, während *eine hübsche Lehrerin* zwar auch eine Zweiwortkombination darstellt, aber doch eher den Status einer freien Wortverbindung hat. Vergleicht man dann *hübsche Lehrerin* beispielsweise mit der Wortverbindung *strenge Lehrerin*, wird deutlich, daß es ganz unterschiedliche Grade an Spezifik bzw. Idiomatizität geben kann, vom Idiom *hübsches Sümmchen* über die aus dem Begriffsfeld von *Lehrerin* resultierende spezifische Wortverbindung *strenge Lehrerin* bis zur freien Kombination *hübsche Lehrerin*.

<sup>3</sup> Ein anderes Beispiel aus der IDS-Datenbank ist *blinder Eifer*. Hier wäre *Eifer* die Basis und *blind* der Kollokator.

Sein Verdienst ist es vor allem, linguistisch systematische Vorstellungen zu Kollokationen gleichzeitig immer in Hinblick auf praktische Bedürfnisse entwickelt zu haben. Für unsere Zwecke erscheint vor allem die Klärung des Verhältnisses zwischen freien Wortverbindungen und Kollokationen fundamental zu sein. Hier gibt es durchaus Parallelen zum Signifikanzkriterium Sinclairs, obwohl sich Hausmann eher im Gegensatz zur britischen Schule definiert.

Die Schule des britischen Kontextualismus gilt bekanntlich als Begründerin der Kollokationsforschung. Beginnend mit Firth stellte sie von Anfang an die empirische Basis (Korpora) in den Mittelpunkt und interessierte sich dafür, welche signifikanten Kollokationen in realen Texten wirklich zu finden sind. Dabei entwickelte sie Computermethoden der statistik- und frequenzorientierten Korpusrecherche. Ein wichtiges Kriterium für die Erfassung von Kollokationen ist die statistische Erwartbarkeit des Mit-einandervorkommens von Wortverbindungen im Korpus. Sinclair versteht unter Kollokationen das Vorkommen zweier oder mehrerer Wörter innerhalb eines bestimmten Abstandes im Text :

Collocation is the occurrence of two or more words within a short space of each other in a text. (Sinclair 1991 : 170)

Wesentliche Kategorien sind dabei 'node', 'collocate' und 'span', die Sinclair folgendermaßen definiert :

'node'

The node word in a collocation (q.v.) is the one whose lexical behaviour is under examination. (ebd. : 175)

'collocate'

A word which occurs in close proximity to a word under investigation is called a collocate of it. (ebd. : 170)

'span'

This is the measurement, in words, of the co-text (q.v.) of a word selected for study. A span of - 4, + 4 means that four words on either side of the node (q.v.) word will be taken to its relevant verbal environment. (ebd. : 175)

Erfaßt wird das formale Kombinationsverhalten eines Wortes mit Kontextpartnern in einer bestimmten (abgegrenzten) Umgebung in großen Textmengen. Zunächst werden die entsprechenden Kontextpartner einer Einheit extrahiert. In einem zweiten Schritt wird dann bestimmt, bei welchen Kombinationen es sich um okkasionelle Bildungen und bei welchen es sich um signifikante Kollokationen handelt. Dazu werden weitere statistische Verfahren hinsichtlich des Vorkommens der einzelnen Elemente im Gesamtkorpus angewandt. Als Ergebnis liegen ausgearbeitete Frequenzlisten signifikanter Kollokationen vor.

Die Herangehensweise des britischen Kontextualismus ist vielfach kritisiert worden, vor allem hinsichtlich der Methoden der Frequenzanalyse. Ihm wird zum einen vorgeworfen, jene Kollokationen nicht erfassen zu können, die nicht frequent sind. Dies meint unter anderem Hausmann, wenn er davon

abráť, die Unterscheidung zwischen einfachen Kookkurrenzen und Kollokationen allein an die Frequenz zu binden, da viele Kollokationen zwar nicht so häufig vorkommen, aber doch verfügbar seien (Hausmann 1985 : 124)<sup>4</sup>. Dazu ist zu bemerken, daß nachzuprüfen wäre, wieviele dieser Kollokationen bei ausreichend großen Korpora wirklich durch das Raster fallen. Hier bietet sich natürlich auch immer an, den Kenntnisstand einzubeziehen, den die Experten über Kollokationen bereits haben. Die Resultate sollten also immer auch anhand bereits kodifizierter Kollokationen in allgemeinen Wörterbüchern und anhand bestehender Kollokationslisten, — wörterbücher und -typologien überprüft werden. Zum anderen wird eingewandt, daß bei Frequenzerhebungen freie Kombinationen, die aufgrund ihrer Häufigkeit das Frequenzkriterium erfüllen, als Kollokationen erscheinen, obwohl sie keine sind. Zu diesem Vorwurf ist zu sagen, daß auch in einem solchen Fall die 'ordnende Hand des Linguisten' unabdingbar ist und immer sein wird. Es wird immer sogenannte 'Fehlbelege' geben, deren Zahl man aber in einem ständigen Bearbeitungsprozeß minimieren kann. Je mehr Klarheit über die sprachliche Spezifik einzelner Wortverbindungen entsteht, desto feiner läßt sich auch ein statistisches Verfahren entwickeln. Das Verdienst des britischen Kontextualismus besteht unbestritten darin, Mechanismen des wirklichen Sprachgebrauchs zu untersuchen, also die Art und Weise, in der Wortverbindungen tatsächlich vorkommen<sup>5</sup>. Solche statistischen Verfahren sind notwendig, um über große Textmengen zu recherchieren und erst einmal eine strukturierte Teilmenge häufig vorkommender Kollokationen zu erhalten.

Frequenz und Statistik stellen eine Basis dar, aber die menschliche Bearbeitung, die philologische Kompetenz muß in jeder Verarbeitungsstufe neu eingesetzt werden. Um eine wirklich praktikable Hilfe für die Produktion fremdsprachiger Texte zu erarbeiten, muß man natürlich über die Konstatierung des reinen Miteinandervorkommens von Wörtern hinausgehen. Aber erst wenn typische Vorkommen, Häufigkeiten und Kontexte aus Korpora

4 Auch in der bereits erwähnten IDS-Datenbank zentraler Adjektiv-Substantiv-Verbindungen des Deutschen gibt es solche Fälle. So weist die Wortverbindung *evangelische Kirche* eine der höchsten Trefferzahlen überhaupt aus, die Kollokation *faule Ausrede* ist dagegen nur 4mal belegt. Natürlich besitzt *faule Ausrede* eine ungleich höhere idiomatische Qualität und muß im Gegensatz zur freien Verbindung *evangelische Kirche*, die den Status eines Eigennamens hat, in einer Kollokationsdatenbank berücksichtigt werden. Deshalb ist die linguistische Kompetenz unverzichtbar. Es bedeutet aber andererseits nicht, daß dieser Befund gar nicht erst zu erheben sei.

5 Stubbs verdeutlicht die Potenz, die in solchen Korpusrecherchen liegt, am Beispiel des Wortes 'pflegen', das er in den Korpora des Instituts für deutsche Sprache in Mannheim untersucht hat. Er erhielt eine lange Liste von Wortkombinationen mit *pflegen*, die häufig vorkommen, aber keine Idiome sind. Dabei wurde als erstes deutlich, daß diese Kontextpartner nur aus einigen wenigen Wortfeldern stammen und die Kollokationen nicht nur zwischen einzelnen Lexemen, sondern auch zwischen Lexemen und semantischen Wortklassen bestehen. "Das Lexem PFLEGEN kann sowohl mit RASEN als auch mit KONTAKT kombiniert werden, aber nicht in beliebigen Wortformen. Man spricht z.B. von einem *gepflegten Rasen*, nicht aber von *\*gepflegten Kontakten*" (STUBBS 1997 : 159) Schließlich stellte Stubbs auch textsortenabhängige Restriktionen von Kombinationen fest. All das ist nach seiner Ansicht ohne Korpusrecherche und statistische Erfassung nicht umfassend beschreibbar.

isoliert sind, ist es möglich, über die Qualität einer Kollokation im Sinne Hausmanns zu urteilen. Sprachliche Regularitäten von Kollokationsbildungen lassen sich natürlich mit Hilfe rein linguistischer Kompetenz generieren. Die Entscheidung, ob eine Wortverbindung nur eine freie Wortverbindung ist oder aber eine kollokative oder gar phraseologische Qualität besitzt, läßt sich also bis zu einem gewissen Grade auf diese Art und Weise fällen. Ist aber der tatsächliche Gebrauch von Interesse, dann müssen wiederkehrende Verwendungsweisen in konkreten Texten ermittelt werden. Erst danach läßt sich eigentlich entscheiden, ob eine Kombination typischer als eine andere ist. Wenn Hausmann meint, daß die Erfahrungen mit dem *Trésor de la langue française* und einem Korpus von ca. 90 Millionen tokens gezeigt haben, daß auch dann immer noch die linguistische Analyse und nicht die Frequenzerhebung den Unterschied zwischen spezifischen und unspezifischen Kollokationen herstelle, ist dem voll zuzustimmen (Hausmann 1985: 126). Dies bedeutet doch aber nicht, nützliche Methoden wie die der britischen Schule nicht verwenden zu können. Man muß u.E. nur genau bestimmen, was man mit einer bestimmten Methode erklären kann und was nicht.

Bevor wir unsere Vorstellungen einer zweisprachigen Kollokationsdatenbank auf der Basis paralleler und vergleichbarer Korpora präzisieren, wollen wir nun einen integrativen Arbeitsbegriff vorgeschlagen. Unter 'Kollokation' verstehen wir typische und wiederkehrende Zweiwortkombinationen im Kontext, d.h. die Verbindung zweier Kerneinheiten und ihrer Kontextpartner (andere Satzglieder oder einbettende Phrasen). Die beiden Bestandteile der Kollokation (Basis und Kollokator bzw. node und collocate) bilden dabei den Kern. Sie sind aber nicht nur als alleinige unmittelbare Kombination von Interesse, sondern auch durch ihre Einbettung in Phrasen, Sätze und Sequenzen. Damit wären nicht nur Wortkombinationen wie *guter Ruf* bzw. *einen guten Ruf genießen* Kollokationen, sondern auch die Realisierung dieser Wortverbindung in der Phrase *Er war besser als sein Ruf*. Der Konstruktion *Der gute Mensch hörte einen Ruf* ist dagegen kein Kollokationsstatus zuzusprechen, obwohl ihre Elemente in der Sinclairschen Kollokationsspanne — hier von jeweils vier Wörtern vor der Basis (*Ruf*) — liegen. Hier muß dann genau die von Hausmann geforderte linguistische Analyse ansetzen, indem zu prüfen ist, inwieweit es Unterschiede im Grad der Kombinierbarkeit und Veränderbarkeit von Kollokationspartnern gibt. Bei der Kollokationsrecherche in den IDS-Korpora konnten beispielsweise sehr viele unterschiedliche sprachliche Realisierungsformen für die Verbindung zwischen dem Substantiv *Kampf* und dem Adjektiv *aktiv* (*aktiver Kampf*, *aktiv im Kampf*, *aktiv am Kampf* usw.) ermittelt werden, die Wortverbindung *aktive Mitarbeit* kommt aber immer nur in ein und derselben Realisierung vor. Ähnliches gilt für Kollokationen wie *erheblicher Teil* oder *in erheblichem Maße*, die im Gegensatz zu *erhebliche Bedeutung* und *erhebliche Schwierigkeiten* nur in dieser unmittelbaren Zweiwortverbindung vorkommen und somit nicht veränderbar scheinen. Der Grad der Veränderungsmöglichkeiten einer Wortverbindung könnte mögli-

cherweise ein Ansatzpunkt für eine zu erarbeitende Kollokationssystematik (eine Skala der Kollokabilität) sein. Je mehr also eine Wortverbindung auch durch andere sprachliche Formen ausgedrückt werden kann, desto weniger Selektionsbeschränkungen wäre sie unterworfen, desto freier würden also ihr Status und ihre Möglichkeit sein, auch mit anderen Kontextpartnern zu kollokieren. Diese Varianten nicht nur aus dem System heraus zu konstruieren, sondern sie anhand konkreter Sprachdaten erheben und nachweisen zu können, darin liegen die großen Möglichkeiten korpusbasierten Arbeitens. Man sollte sich trotzdem zunächst in den zweisprachigen Listen auf klassische Zweiwortkombinationen (z.B. Adjektiv-Nomen oder Verb-Nomen) konzentrieren, um nicht in der Komplexität der Belege 'unterzugehen'. Diese extrahierten Zweiwortkombinationen und ihre Übersetzungsäquivalente stellen den Kern dar, von dem aus aber eine weitergehende Systematik entwickelt werden muß, in der Kollokationen und ihre Äquivalente in der anderen Sprache in ihrem kontextuellen Umfeld erfaßt und systematisiert angeboten werden können.

### **3 Eine Kollokationsdatenbank für das Sprachenpaar Deutsch-Französisch — erste konzeptionelle Vorstellungen**

#### **3.1 Kollokationen in traditionellen und in elektronischen Wörterbüchern**

Die existierenden monolingualen und bilingualen Wörterbücher enthalten in ihren Artikeln eine Vielzahl von Kollokationen, idiomatischen Wendungen und Wortverbindungen, jedoch in der Regel ohne eine syntaktische und semantische Systematik. Darüber hinaus stehen die Lexikographen naturgemäß immer vor großen Problemen hinsichtlich der Zugriffsmöglichkeiten und der Entscheidung für eine Artikelstruktur, die auf kleinem Raum möglichst viele — schnell nachvollziehbare — Informationen beinhalten sollte.

Ein Problem besteht zunächst in der Auswahl der Belege. Die bisherigen lexikographischen Beleglisten haben nach wie vor einen intuitiven und eher zufälligen Charakter; das heißt, die Entscheidung für einen Wörterbucheintrag hängt zu einem großen Teil von der Bewertung des einzelnen Lexikographen ab. Er entscheidet aufgrund seiner Sprachkompetenz, ob er eine Variante für einschlägig und relevant hält<sup>6</sup>. Stubbs problematisiert diese

<sup>6</sup> Es gibt mittlerweile eine wachsende Zahl von Spezialwörterbüchern, die durchaus auch auf Korpora basieren. Aber auch bei jenen lexikographischen Projekten, die bereits auf großen Korpora aufbauen, wird schließlich eingeräumt, diese in letzter Instanz doch nur zur Validierung selbst ausgewählter Belege genutzt zu haben. Die eigene Sprachkompetenz und -erfahrung bleibt Maßstab für Relevanzsetzungen und für im Wörterbuch zu fixierende Kontextangebote (vgl. u.a. ZIMMER 1990, SCHEMANN 1994). Es wurden keine automatisierten statistischen Frequenzanalysen durchgeführt, die zu einer ungleich größeren, manuell nicht mehr zu bewältigenden Belegmenge führen. Für das Sprachenpaar "Französisch-



intuitiven Zuordnungen am bereits erwähnten Beispiel der Kombinationen mit dem Verb *pflegen*. Er habe durch Korpusrecherche eine ungleich größere Kombinationsvielfalt isolieren können, als das durch die Kompetenz des einzelnen Forschers möglich wäre :

Zweitens kann es auch sein, daß Sie glauben, Sie hätten solche Ergebnisse selbst herausgefunden, aus Ihrem eigenen Sprachgefühl, wenn Sie nur einige Minuten darüber nachgedacht hätten. Hier kann ich nur sagen — [...] — daß Muttersprachler zwar oft in der Lage sind, einige Beispiele solcher Muster intuitiv zu finden. Und solche Beispiele sind oft richtig. Sie sind aber nicht in der Lage, solche Muster systematisch und gründlich zu belegen. (Stubbs 1997 : 159)

Überwunden werden kann diese intuitive Komponente nur durch Frequenzanalysen auf der Basis großer, kohärenter Korpora. Im vorgeschlagenen Fall handelt es sich um parallele Korpora, die Texte und ihre Übersetzungen in der jeweils anderen Sprache enthalten, und um vergleichbare Korpora mit analogen Texten eines Kommunikationsbereichs, zu einem identischen Thema usw. (vgl. 3.2.). Ein weiteres Problem traditioneller Wörterbücher ist die Strukturierung der Einträge. Wo soll zum Beispiel die Kollokation verzeichnet sein, unter Basis oder Kollokator (vgl. dazu Hausmann 1985 : 118 ff)? Das traditionelle, gedruckte zweisprachige Wörterbuch ist und bleibt auf das alphabetisch geordnete Einzelwort bezogen. Je mehr sich moderne Lexikographen bemühen, die Idiomatik zu berücksichtigen, desto unübersichtlicher werden notgedrungen die Worteinträge sein. Lexikographen versuchen, diese Klippe zu überwinden, indem sie ein Netz von Querverweisen erstellen, was natürlich zu immer größeren Rezeptionsproblemen führt, je komplexer die Kontextangaben werden.

### 3.2 Parallele und vergleichbare Korpora

Der auch in der Korpuslinguistik noch relativ neue Ansatz, zwei- oder mehrsprachige Übersetzungswerkzeuge auf der Basis paralleler und vergleichbarer Korpora (vgl. dazu Teubert 1996 a und 1996 b) zu entwickeln, ist vor allem durch die eben dargestellten Hintergründe motiviert. Parallelkorpora bestehen aus Originaltexten und deren Übersetzungen bzw. äquivalenten Übersetzungstexten in verschiedenen Sprachen. Alle Texte werden durch automatische Verfahren miteinander in Beziehung gesetzt und können bis in ihre kleinsten Elemente hinein miteinander verglichen werden. Der Vorteil dieser Parallelkorpora besteht darin, daß sie eine Vielzahl an Übersetzungsvarianten bieten, die weder ein einzelner Lexikograph noch ein einzelner Übersetzer in dieser Menge abrufen kann. Sie enthalten das implizite Sprachwissen des Übersetzers, das nicht nur mehr umfaßt, sondern anders organisiert ist als die expliziten Informationen, die in bilingualen Wörterbüchern enthalten sind. Dieses implizite Übersetzungswissen muß durch geeignete Verfahren verfügbar gemacht werden. Man wird bei diesen

Deutsch/Deutsch-Französisch" seien hier vor allem ZIMMER 1990, BARDOSI et al. 1992, ILGENFRITZ et al. 1993, SCHIEMANN 1994 erwähnt.

Übersetzungskorpora jedoch immer eine gewisse subjektive Komponente finden, die in den Eigenheiten einzelner Übersetzungsstile begründet ist. Daher ist es notwendig, die erhobenen Befunde anhand von analogen (vergleichbaren) Korpora in der jeweiligen Einzelsprache zu validieren. Solche vergleichbaren Korpora enthalten Texte, die in Größe und Aufbau gleich strukturiert, aber keine Übersetzungen voneinander sind. Man kann z.B. Texte in verschiedensprachigen Tageszeitungen gegenüberstellen, wenn sie sich in ausgewählten textinternen und -externen Parametern, beispielsweise hinsichtlich der Thematik entsprechen (z.B. Texte in der deutschen und französischen Presse zur Einführung der einheitlichen europäischen Währung).

Die empirische Basis für eine zu erarbeitende Kollokationsdatenbank soll u.a. folgendes Parallelkorpus bilden: Seit 1995 erscheint die französische Wochenzeitung *Le Monde diplomatique* gleichzeitig in verschiedenen Sprachen, darunter auch in einer deutschen Übersetzung. Es handelt sich um ein exemplarisches Parallelkorpus aus der unmittelbaren Sprach- und Übersetzungspraxis mit ca. 1 Millionen Wortformen pro Sprache. Für die Validierung der Ergebnisse soll zum einen ein Korpus mit Übersetzungen vom Deutschen ins Französische, zum anderen sollen vergleichbare Texte von jeweils einer anderen Tages- bzw. Wochenzeitung (z.B. *Le Monde* und *Frankfurter Rundschau*) hinzugezogen werden. Mit der Entscheidung für diese Zeitungskorpora wird auch eine Eingrenzung des zu erfassenden Kollokationsvorkommens und der entsprechenden Übersetzungsäquivalente vorgenommen, was für diese Erprobungsphase sinnvoll erscheint. Zeitungskorpora sind in mehrfacher Hinsicht nutzbringend: Zum einen handelt es sich um einen abgegrenzten, gut strukturierten Kommunikationsbereich, der vergleichsweise klar definierte und in beiden Ländern analoge Textsorten (Genres) aufweist. Zum anderen bildet der Bereich der öffentlichen Kommunikation sowohl den politisch-gesellschaftlichen Diskurs allgemein ab als auch große Teile der Alltagssprache und ermöglicht somit verallgemeinerbare Schlußfolgerungen für den Gebrauch von Kollokationen.

### 3.3 Was bedeutet eine zweisprachige Kollokationsdatenbank ?

Für die Erstellung einer schnell abrufbaren Datenbank relevanter Übersetzungsäquivalente aller gebräuchlichen Kollokationen im Französischen und Deutschen sind drei Schritte nötig: umfangreiche Kollokationsrecherchen in parallelen und vergleichbaren Korpora; die Erarbeitung einer anwendungsbezogenen Kollokationssystematik und schließlich die Aufbereitung dieser Ergebnisse in einer lexikographischen Datenbank.

Zunächst müssen die signifikanten Kollokationen in der jeweiligen Einzelsprache aus dem Korpus extrahiert werden. Für die Extraktion von Kollokationen gibt es bereits automatische tools (u.a. das IDS-eigene Kollokationsrecherchesystem unter COSMAS). Für den Sprachvergleich sind Verfahren zur Parallelisierung von Sätzen, Sequenzen und Texten (lexical alignment)

von Bedeutung. Einzusetzen sind darüber hinaus elektronische Abfrageprozeduren in existierenden Wörterbüchern (lexicon lookup)<sup>7</sup>. Aus den im Ergebnis dieser ersten Kollokationsextraktionen entstandenen vollständigen Frequenzlisten aller Kombinationsmöglichkeiten von Zweiworteinheiten und ihrer kontextuellen Einbettungen lassen sich diejenigen Verbindungen mit der höchsten Vorkommensfrequenz bestimmen. Danach ist zu entscheiden, welche Kollokationsvorkommen signifikant sind und welche eher marginal. Darüber hinaus wird es aber auch Kollokationen geben, die zwar nicht das Kriterium größter Vorkommenshäufigkeit erfüllen, denen aber aus ganz bestimmten Gründen dennoch eine Relevanz für den jeweiligen Text, die jeweilige Textsorte oder Sprache generell zugesprochen werden muß. Dafür Relevanzkriterien zu entwickeln, scheint ein lohnendes Ziel für eine linguistische Analyse zu sein. Danach können die französischen und deutschen Befunde miteinander verglichen werden, um herauszufinden, ob es invariante Kernbereiche von Kollokationen gibt und wo die Variablen liegen. Es muß also gefragt werden, welche Kollokationen in beiden Sprachen ähnlich zentral sind und welche im Deutschen charakteristischen Kollokationen im Französischen keine zentrale Rolle spielen (bzw. umgekehrt). Damit ist es dann möglich, die jeweiligen Übersetzungsäquivalente gegenseitig zuzuordnen, wodurch im Detail geklärt werden kann, inwieweit Kollokationen der einen Sprache auch mit Kollokationen in der anderen Sprache übersetzt wurden<sup>8</sup>.

Der nächste wichtige Schritt ist die automatische Erstellung und Systematisierung typischer Kontextuierungen<sup>9</sup> von Kollokationen, ein nun wirk-

7 Vgl. dazu ausführlicher STEYER/TEUBERT 1998.

8 Erste Stichprobenanalysen lassen vermuten, daß Kollokationen in der einen Sprache nicht automatisch mit Kollokationen in der anderen Sprache übersetzt werden, daß vielmehr ein großer Prozentsatz "mit anderen Worten" paraphrasiert wird. Diese Übersetzungsäquivalente können dann natürlich nicht im Wörterbuch verzeichnet sein. Dazu sei ein Beispiel aus einer anderen Fallstudie angeführt, bei der Kollokationsvorkommen in einer deutschen und französischen Übersetzung von Platons philosophischer Abhandlung "Der Staat" analysiert wurden. Auch wenn es um den Vergleich paralleler Übersetzungen aus einer Drittsprache ging, lohnt sich ein Blick auf die Ergebnisse: Signifikant vorkommende Adjektiv-Substantiv-Verbindungen in der deutschen Fassung waren z.B. *breite Masse* und *wahres Sein*. Die französischen Parallelformen in den jeweiligen Textabschnitten sind ganz unterschiedlicher Natur. So bietet der französische Text an der Stelle, an der im deutschen Text *breite Masse* vorkommt, folgende Äquivalente an: *la masse*, *la masse de gens*, *la masse du peuple* und *la plupart des gens*. Keines dieser Äquivalente steht im Wörterbuch als mögliche Entsprechung. Für die in der deutschen Fassung häufig vorkommende Kollokation *wahres Sein* findet sich im Korpus bis auf eine Ausnahme (*l'être*) als französisches Übersetzungspendant *qui est réel*, eine Verbalphrase, die im zweisprachigen Wörterbuch nicht unter (*wahres*) *Sein* verzeichnet sein kann. Dort findet man für die philosophischen Terminus *Sein* französische Entsprechungen wie *l'être*, *l'essence* oder *l'existence*. Der Übersetzer hat diese Entsprechungen jedoch nicht verwendet, sondern eine andere Verbalphrase. Dies könnte möglicherweise mit dem Stil der französischen Übersetzung zusammenhängen, die im Ganzen Alltagssprachlicher angelegt ist als die dem philosophischen Stil verhaftete deutsche Übersetzung (vgl. hierzu ausführlich STEYER/TEUBERT 1998). Ohne Einbeziehung der textuellen Faktoren läßt sich die Übersetzung schwer nachvollziehen. Das Wörterbuch bietet in diesem Fall wenig Erklärungshilfe.

9 'Kontext' ist hier als ein dynamisches Konzept zu verstehen. Zunächst werden mit 'Kontext' in Hinblick auf automatische Extraktionen die unmittelbaren sprachlichen Partner

lich völlig neuartiges Verfahren, das jedoch für praktikable Übersetzungsangebote als absolut dringlich erscheint. Die als relevant bestimmten Kollokationen werden also nicht nur im jeweiligen Korpus recherchiert, sondern mit ihren Kontexten und den Parallelvarianten in der anderen Sprache kontrastiert. Beispielsweise können alle Kontextbelege einer signifikanten Kollokation im Deutschen aufgelistet und gleichzeitig die äquivalenten französischen Übersetzungsbelege zugeordnet werden. Damit läßt sich zeigen, auf welcher unterschiedlichen Art und Weise ein und dieselbe Ausgangseinheit in der Quellsprache in den Übersetzungen in welchen Kontexten expliziert wird.

Im Mittelpunkt aller Analysen steht im Sinne der eben entwickelten Argumentation immer die Frage, ob es Übersetzungsäquivalente gibt, die das klassische Wörterbuch nicht anbietet. Ist dies der Fall, so stellt sich in der weiteren Analyse die Frage, ob diese nur den Status von einzeltext- bzw. einzelsprecherbezogenen Ad-hoc-Bildungen haben oder ob sie bereits Usus sind. Es gilt also, Standardverwendungen herauszufinden, die so nicht im Wörterbuch fixiert sind. Statistische Voruntersuchungen haben ergeben, daß ein großer Teil von Übersetzungseinheiten in solchen Parallelkorpora nicht im zweisprachigen Wörterbuch erfaßt ist<sup>10</sup>. Es geht also vorrangig um die Darstellung von Kollokationen, die auf Grund ihrer Vorkommenshäufigkeit oder ihrer textuellen Relevanz durchaus signifikant sind, aber mit bisherigen lexikographischen Hilfsmitteln nicht erfaßt und beschrieben wurden. Solche Abweichungen, die ja möglicherweise gar keine sind, sondern ebenso Regularitäten und Standardvarianten darstellen, können nur mit Hilfe der vorgeschlagenen listen- und regelbasierten Recherchen und Analysen erfaßt werden. Durch kontextuelle Gegenüberstellungen läßt sich weiterhin analysieren, inwieweit sich auch die entsprechenden Einbettungsstrukturen wiederholen. Man kann fragen: Gibt es spezifische sprachstrukturelle und kommunikative Umgebungen für bestimmte Kollokationen? Sind diese Umgebungen generalisierbar für bestimmte textuelle Konstellationen, Textsorten, Stilebenen usw.? Diese Fragen betreffen praktische Bedürfnisse der Übersetzungsgararbeit in elementarer Weise: Die elektronische Kollokationsdatenbank wird nicht nur eine systematisierte Ordnung häufig verwendeter

einer Kollokation (unterschiedliche Satzglieder bzw. Wortarten) erfaßt. Von der systematisierten Erfassung dieser Umgebungen kann man dann zu einer Systematik der sequentiellen Einbettungen und der Kontexte in bezug auf Textsorten, Kommunikationssituationen, Adressaten usw. kommen. Damit können dem Übersetzer schließlich gebräuchliche Verwendungskontexte angeboten werden. Auch ein Beispiel aus der Plato-Analyse: Für die im Korpus häufig vorkommene Wortkombination *mit vollem Recht* bietet das Wörterbuch tatsächlich einige Übersetzungsmöglichkeiten an (*à bon droit, à juste titre, de plein droit, par droit et raison*). Der französische Plato-Übersetzer verwendet jedoch keine dieser Phrasen, sondern Fügungen bzw. lexikalische Einheiten wie *tu disais vrai* oder *justement*. Diese Übersetzungen werden erst dann nachvollziehbar, wenn man den illokutiv-kommunikativen Kontext dieser Ausdrücke kennt. In unserem Fall handelt es sich bei der Äußerung mit vollem Recht um einen reinen Sprechakt des Bestätigens in einem philosophischen Dialog. Solche Bestätigungshandlungen können im Französischen mit ganz anderen sprachlichen Formulierungen ausgedrückt werden, die im Wortfeld von *Recht* oder *volles Recht* nicht zu finden wären.

<sup>10</sup> Vgl. dazu GEYKEN 1997.

Kollokationen und ihrer Übersetzungsäquivalente enthalten, sondern auch abrufbare, systematisierte Kontextmuster<sup>11</sup>.

### 3.4 Kollokationen in Paralleltexten — einige Beispiele

Da sich das Projekt für eine zweisprachige Kollokationsdatenbank erst in einer konzeptionellen Pilotphase befindet, können natürlich an dieser Stelle noch keine Resultate vorgestellt werden, z.B. die Validierung erster Annahmen anhand großer Textmengen<sup>12</sup>. Die Beispiele haben somit lediglich Demonstrationscharakter. Grundlage für diese ersten empirischen Stichprobenanalysen bilden zwei Ausgaben der *Le Monde diplomatique* und ihre jeweiligen deutschen Ausgaben als Beilage der *tageszeitung*. Eine erste Kollokationsrecherche in den Paralleltexten hat ganz unterschiedliche Resultate erbracht: Für eine bestimmte Menge französischer Kollokationen lassen sich äquivalente deutsche Entsprechungen nachweisen. Diese 1:1-Entsprechungen waren für unsere analytischen Zwecke dann nicht von Interesse, wenn sie bereits in einschlägigen Wörterbüchern kodifiziert waren. Es bleibt aber eine nicht unbeachtliche Restmenge nichtäquivalenter Ersetzungen. Die Grundfrage ist auch hier, ob es sich bei diesen Ersetzungen um rein okkasionelle, individuelle Bildungen handelt oder ob sie bereits usuell sind.

Das Beispiel (1) zeigt einen solchen Fall, bei dem die Transformation allem Anschein nach individueller Natur ist; aber wirklichen Aufschluß kann letztlich nur die Validierung im Korpus bringen:

<i>Le Monde diplomatique</i>	
französische Ausgabe, 17.10.1997, S. 3 (1)	deutsche Ausgabe, 17.10.1997, S. 3
Französisch	Deutsch
<i>Cette villa Susini avait été un lieu célèbre, dont le nom est passé à la postérité. Durant la lutte de libération nationale, elle avait été un centre de torture où de nombreux résistants et résistantes trouvèrent la mort.</i>	<i>Diese Villa Susini war ein berühmter, berühmter Ort, dessen Name der Nachwelt noch bekannt ist. Während des Befreiungskampfes war die Villa ein Folterzentrum gewesen, wo zahlreiche Widerstandskämpfer und -kämpferinnen den Tod gefunden hatten.</i>

Die Einträge für die deutsche Entsprechung des Adjektivs *célèbre* in den hier betrachteten zweisprachigen Wörterbüchern lauten alle in einem ähnlichen Sinne *berühmt, weithin bekannt*; auf eine Person bezogen auch *gefeiert*; das zu diesem Feld gehörende Verb *célébrer* wird mit *feiern, festlich begehen, Gedenkfeier veranstalten, rühmen, preisen* usw. übersetzt. In

<sup>11</sup> Zu Relevanz der Kontexte vgl. auch Teubert in diesem Heft.

<sup>12</sup> Wie diese Recherche an einem, allerdings sehr speziellen Parallelkorpus aussehen kann und welche Ergebnisse zutage treten können, haben wir an anderer Stelle vorgeführt (vgl. STEYER/TEUBERT 1998).

keiner Bedeutungserklärung ist jedoch die negative Konnotation enthalten, die dieses Adjektiv im vorliegenden Kontext eindeutig aufweist. Man kann nun darüber spekulieren, warum der Verfasser des französischen Originaltextes nicht einen adäquateren französischen Ausdruck verwendet hat, etwa das vom *Petit Robert* angebotene *tristement célèbre* oder das für Ortskennzeichnungen mögliche *mal famé*. Der Übersetzer hat jedenfalls versucht, die seiner Meinung nach vom Kontext zwingend gebotene Bewertung durch die lexikalisierte Kollokation *berühmt-berüchtigt* deutlich zu machen. Es bleibt aber die Frage offen, ob es sich bei der Verwendung von *célèbre* mit einer negativen Konnotation um eine individuelle sprachliche Fehlleistung handelt, die vom Übersetzer korrigiert wurde, oder um eine vom Wörterbuch nicht verzeichnete Verwendungsweise NEG *célèbre*, die dann im Deutschen die Kollokationsentsprechung *berühmt-berüchtigt* hätte. Aufschluß darüber kann nur die Recherche an großen Datenmengen geben<sup>13</sup>.

Zur Veranschaulichung der geplanten Herangehensweise soll nun ein ganzer Textabschnitt und seine Übersetzung betrachtet werden. Es handelt sich bei Beispiel (2) um einen Artikel der *Le Monde diplomatique* unter dem Titel: "Sur la télévision" und seine Übersetzung in der deutschsprachigen Ausgabe ("Bourdieu Fernsehkritik"). Wenn man Paralleltexte verwendet, ist es bereits auf automatischem Wege möglich, diese Parallelisierung auch optisch sichtbar zu machen:

<i>Le Monde diplomatique</i>	
französische Ausgabe, 14.2.1997, S. 31 (2)	deutsche Ausgabe, 14.2.1997, S. 11
Französisch	Deutsch
<p><i>Si "la télévision fait courir un danger à la vie et à la démocratie", c'est qu'à l'exposé d'un "discours articulé" elle préféra toujours "la parole des penseurs qui sont censés penser à vitesse accélérée".</i></p> <p><i>Dans son nouveau livre (1), Pierre Bourdieu choisit de ne pas s'attarder sur les contraintes essentielles — mais connues — qui pèsent sur la profession de journaliste, nourrissant quelques-uns de ses tropismes (idéologie de marché, réseaux de connivence, révérence pour le pouvoir, tentations de la complaisance, du nombrilisme et de l'argent).</i></p>	<p><i>Wenn "das Fernsehen eine Bedrohung für die Demokratie darstellt", dann, weil es dem "geschwätzigen Denkerer, die angeblich mit erhöhter Geschwindigkeit denken", den Vorzug gibt gegenüber der "durchdachten Rede".</i></p> <p><i>Pierre Bourdieu hält sich in seinem neuen Buch (1) nicht lange mit den Zwängen auf, denen Journalisten unterworfen sind und die einigen ihrer professionsbedingten Reflexen Vorschub leisten (Marktideologie, politische Feigheit, gepaart mit Komplizenschaft, Machthörigkeit, Selbstgefälligkeit, Faible für große Namen und Geld).</i></p>

13 Natürlich kann ein Muttersprachler aus seinem Sprachgefühl urteilen, eine Korpusrecherche bringt aber gesichertere Erkenntnis über diese Verwendungen und bietet zudem die Möglichkeit, weitere Verwendungskontexte von *célèbre* zu dokumentieren.

Auf der Basis dieser Parallelisierung kann man nun alle Adjektiv-Nomen-Verbindungen und alle Verb-Nomen-Verbindungen extrahieren und in der Form, in der sie in den Texten vorkommen, gegenüberstellen :

Französisch	Deutsch
<i>fait courir un danger</i>	<i>eine Bedrohung darstellt</i>
—	<i>geschwätziges Denken</i>
<i>un discours articulé</i>	<i>der durchdachten Rede</i>
<i>(préfère)</i>	<i>den Vorzug gibt</i>
<i>sont censés penser à vitesse accélérée</i>	<i>angeblich mit erhöhter Geschwindigkeit denken</i>
<i>dans son nouveau livre</i>	<i>in seinem neuen Buch</i>
<i>les contraintes essentielles</i>	<i>den Zwängen</i>
<i>les contraintes pèsent sur</i>	<i>den Zwängen unterworfen</i>
<i>(nourrissant)</i>	<i>Vorschub leisten</i>

Danach ist es möglich, einzelne Transformationsprozesse näher zu bestimmen. In unserem Beispiel lassen sich u.a. folgende Unterscheidungen treffen :

- a) 1:1-Entsprechungen bei den Zweiwortkombinationen ohne Kollokationsstatus :

*[dans son] nouveau livre vs. [in seinem] neuen Buch*

- b) Ersetzungen einer Kollokation durch eine Zweiwortkombination, die aber anders lexikalisiert ist (Dieser Fall tritt in den Texten sehr häufig auf.) :

*fait courir un danger vs. eine Bedrohung darstellen:*

*Courir un/le danger* ist eine Kollokation, die im Wörterbuch verzeichnet ist (Sachs/Villatte : "danger [...] Gefahr [...] courir le~de (+ inf) Gefahr laufen [...]") Dagegen wird die freie deutsche Wortkombination *eine Bedrohung darstellen* gesetzt.

*les contraintes pèsent sur vs. den Zwängen unterworfen :*

Zunächst ist festzustellen, daß es für die Phrase *les contraintes pèsent sur* weder unter *contrainte* noch unter *peser* einen entsprechenden Eintrag in den Wörterbüchern gibt. Es handelt sich daher entweder um eine freie, beliebige Wortverbindung oder um eine zwar verwendete, aber nicht kodifizierte Kollokation. Der Übersetzer ersetzt eine freie aktivische Wortverbindung durch die passivische Kollokation *Zwängen unterworfen sein*. Er hätte auch "wörtlicher" übersetzen können: ["... hält sich nicht lange mit den Zwängen auf, die auf dem Journalistenberuf lasten"). Hier ließe sich durch Korpuserhebungen nachprüfen, welche weiteren französischen Äquivalente für diese Kollokation in welcher Frequenz vorkommen. Daraus könnte man dann schlußfolgern, ob durchaus mögliche französische äquivalente Kollokationen — z.B. *être soumis à des contraintes* — wirklich

häufig verwendet werden. Es kann aber durchaus auch möglich sein, daß die verwendeten französischen Entsprechungen ganz unterschiedlich sind, also die lexikalische Einheit *contrainte* freier kombinierbar ist. In diesem Fall bietet die Auswertung des Parallelkorpus die Möglichkeit, eine systematisierte Liste häufig verwendeter typischer Übersetzungsäquivalente mit ihren Kontexten zu erarbeiten, aus denen der Übersetzer bzw. Produzent eines fremdsprachigen Textes auswählen kann.

- c) Ersetzungen, bei denen nur ein Äquivalent in einer der beiden Sprachen eine Mehrworteinheit bzw. eine Kollokation ist (vgl. auch die weiter oben diskutierte Ersetzung von *célèbre* durch *berühmt-berüchtigt*) :

(*préféra*) vs. *den Vorzug gibt*

[Ein Verb wird durch eine Kollokation ersetzt]

*les contraintes essentielles* vs. *die Zwänge*

[Weglassen des Adjektiv *essentielle*]

(*nourrissant*) vs. *Vorschub leisten*

[Ein Verb wird durch eine Kollokation ersetzt]

- d) Ersetzungen, die schon aufgrund der Originalformulierungen sprecherbezogene Ad-Hoc-Bildungen und daher kaum verallgemeinerbar sind :

*sont censés penser à  
vitesse accélérée*

*angeblich mit erhöhter Geschwindigkeit  
denken*

- e) Ersetzungen, die aus dem Kontext ableitbar und durchaus gebräuchlich sind :

*un discours articulé* vs. *durchdachte Rede*

Zunächst wird die Kollokation *discours articulé* daraufhin überprüft, welche Bedeutungsvarianten die Wörterbücher anbieten. Dem französischen Nomen *discours* entspricht — von Nebenbedeutungen abgesehen — in den meisten Fällen das deutsche Nomen *Rede*. Zum Adjektiv *articulé* gibt es differenziertere Angaben :

Sachs/Villatte :

"articulé[...] 4. *fig. Rede, Abhandlung* gegliedert; *Gedanken* ~ l'un sur l'autre miteinander verbunden, verknüpft [...]"

"articuler[...] 1. *phon* artikulieren; deutlich aussprechen [...] s~ [...] 6. *fig. Gedanken, (Rede) Teile, Kapitel* miteinander verbunden, verknüpft sein; ineinander gefügt sein [...]"

Weis/Mattutat :

"articulé, -e *a* gegliedert; [...] *fig* klar, deutlich vernehmbar [...]"

"articuler [...] gliedern; inea. = fügen; [...] *fig* aus = sprechen, behaupten, sagen [...]"

Dabei fällt auf, daß man bei der Umkehrprobe, also bei der Suche nach französischen Äquivalenten für *durchdacht*, in keinem Fall *articulé* als eine mögliche Variante findet :



Sachs/Villatte :

"durchdacht [...] *adjt. Plan etc.* mûrement réfléchi; ~denken *v/t* [...] examiner à fond; réfléchir profondément à; méditer, ruminer"

Weis/Matutut :

"[...] durchdenken *tr* approfondir, examiner à fond; méditer, ruminer → *gut durchdacht* bien réfléchi"

Die in den Wörterbüchern angebotenen Äquivalente von *articulé* beziehen sich fast durchweg auf technisch-rhetorische Wertzuschreibungen : Eine Rede ist gut aufgebaut, gut gegliedert etc. Im deutschen Text wird dafür jedoch *durchdacht* verwendet (vgl. Wörterbucheintrag zu *durchdenken* im Duden Deutsches Universalwörterbuch : "*durchdenken* [...] *in Gedanken von Anfang bis Ende durchgehen, [...] vollständig, in allen Einzelheiten, hinsichtlich der Möglichkeiten u. Konsequenzen überdenken* [...]"; Wahrig : "[...] *genau nachdenken über, gründlich überlegen* [...]"). Es geht ja in der Tat in diesem Textabschnitt um die abstraktere Bedeutung von *Rede* (im Sinne von Rede-/Sprechweise) und um die Qualitäten dieser Redeweise. Der Übersetzer hat sich also unter Berücksichtigung des Textzusammenhangs für eine nicht im Wörterbuch verzeichnete Variante entschieden, die für das Deutsche aber adäquater erscheint. Es ließe sich nun anhand des Parallelkorpus nachprüfen, inwieweit auch an anderen Stellen Wortverbindungen, in denen *articulé* vorkommt, mit *durchdacht* übersetzt wurden. Treten diese Übersetzungen mit hoher Frequenz auf, ist zu fragen, ob man diese Kombinationen mit *articulé* im Sinne von *durchdacht* nicht ins Wörterbuch aufnehmen sollte.

#### 4 Zusammenfassung

Im vorliegenden Beitrag ging es zum einen darum, bestehende Kollokationskonzeptionen in einer ersten Annäherung für die Zwecke einer Datenbank zu prüfen. Es wurde ein integrativer Vorschlag für ein Kollokationskonzept unterbreitet, das sowohl die Methoden des britischen Kontextualismus als auch stärker philologisch-systematisierende Kollokationsansätze berücksichtigt. Zum anderen konnte — wenn auch nur sehr punktuell — angedeutet werden, wie eine bilinguale Recherche und die Aufbereitung relevanter Übersetzungseinheiten für eine zweisprachige Kollokationsdatenbank auf der Basis von Paralleltexten aussehen könnte. Dabei ist hervorgehoben worden, welchen Gewinn diese korpusbasierte und kontextorientierte Herangehensweise sowohl für die linguistische Forschung als auch für die praktische Übersetzungsarbeit haben kann.

## Bibliographie

- BAHNS Jens, 1996. *Kollokationen als lexikographisches Problem. Eine Analyse allgemeiner und spezieller Lernerwörterbücher des Englischen* (= Lexicographica : Series Maior 74), Tübingen.
- BÁRDOSI Vilmos / ETTINGER Stefan / STÖLTING Cécile, 1992. *Redewendungen Französisch-Deutsch. Thematisches Wörter- und Übungsbuch*, Tübingen.
- BRESSON Daniel / KUBCZAK Jacqueline (eds), 1998. *Abstrakte Nomina. Vorarbeiten zu ihrer Erfassung in einem zweisprachigen syntagmatischen Wörterbuch* (= Studien zur deutschen Sprache 10), Tübingen.
- DUDEN *Deutsches Universalwörterbuch*, 1989. 2., völlig neu bearb. und stark erw. Aufl., Wiss. Rat und Dudenredaktion unter Leitung v. G. Drosdowski (eds), Mannheim, Wien, Zürich.
- GEYKEN Alexander, 1997. "Matching Corpus Translations with Dictionary Senses : Two Case Studies", in *International Journal of Corpus Linguistics*, vol. 2 (1), 1-22.
- HAUSMANN Franz Josef, 1985. "Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels", in *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6.1984*, H. Bergenholtz / J. Mugdan (eds) (= Lexicographica 3), Tübingen, 118-129.
- , 1989. "Le dictionnaire de collocations", in *Wörterbücher. Ein internationales Handbuch zur Lexikographie*, F.J. Hausmann / O. Reichmann / H.E. Wiegand / L. Zgusta (eds), Erster Teilband (= Handbücher zur Sprach- und Kommunikationswissenschaft 5.1), Berlin, New York, 1010-1019.
- , 1993. "Was ist und was soll ein Kontextwörterbuch ?" (Einführung), in *Langenscheidts Kontextwörterbuch Französisch-Deutsch. Ein neues Wörterbuch zum Schreiben, Lernen, Formulieren*, 5. Aufl., P. Ilgenfritz / N. Stephan-Gabinel / G. Schneider, Berlin, München, Wien, Zürich, New York..
- KLEIBER Georges, 1993. *Prototypensemantik. Eine Einführung*, Tübingen.
- LEHR Andrea, 1996. *Kollokationen und maschinenlesbare Korpora. Ein operationales Analysemodell zum Aufbau lexikalischer Netze* (= Reihe Germanistische Linguistik 168), Tübingen.
- NOUVEAU PETIT ROBERT, 1993. *Dictionnaire alphabétique et analogique de la langue française*. Nouvelle édition remaniée et amplifiée sous la direction de J. Reg-Debove et Alain Rey du Petit Robert par Paul Robert, Paris.
- SACHS Karl / VILLATTE Césaire, 1994. *Langenscheidts Großwörterbuch Französisch. Teil I. Französisch-Deutsch*, E. Weis (ed), 8. Aufl. der V. Neubearb. 1979, Berlin, München, Wien, Zürich.
- , 1979. *Langenscheidts Großwörterbuch Französisch. Teil II. Deutsch-Französisch*. W. Gottschalk / G. Bentot (eds). Völlige Neubearb., 13. Aufl. der V. Neubearb. 1968. Mit Nachtrag 1979, Berlin, München, Wien, Zürich.
- SCHEMANN Hans / RAYMOND Alain, 1994. *Pons Idiomatik Deutsch-Französisch. Dictionnaire idiomatique allemand-français*, Stuttgart, Dresden.
- SINCLAIR John, 1991. *Corpus, Concordance, Collocation*, Oxford.
- STEYER Kathrin / TEUBERT Wolfgang, 1998. "Deutsch-Französische Übersetzungsplattform — Ansätze, Methoden, empirische Möglichkeiten", in *Deutsche Sprache*. (erscheint)

- STUBBS Michael, 1997. "Eine Sprache idiomatisch sprechen": Computer, Korpora, Kommunikative Kompetenz und Kultur", in *Norm und Variation*, K. Mattheier (ed) (= forum Angewandte Linguistik 32), Frankfurt a.M., Berlin, Bern, New York, Paris, Wien, 151-167.
- TEUBERT Wolfgang, 1996 a. "Comparable or Parallel Corpora?", in *International Journal of Lexicography*, Vol. 9 No.3. Oxford, 238-276.
- , 1996b. "Deutsch-französische Verständigung. Ein Übersetzungswerkzeug für das 21. Jahrhundert.", in *Sprachreport* 2/96, Mannheim, 7-11.
- VIEHWEGER Dieter (et al.), 1977. *Probleme der semantischen Analyse* (= Studia grammatica XV), Berlin.
- VIEHWEGER Dieter, 1982. "Die Darstellung semantischer Vereinbarkeitsbeziehungen zwischen lexikalischen Elementen im einsprachigen Wörterbuch des Deutschen", in *Wortschatzforschung heute: Aktuelle Probleme der Lexikologie und Lexikographie*, E. Agricola / J. Schildt / D. Viehweger (eds), Leipzig, 23-41.
- WAHRIG Gerhard, 1994. *WAHRIG Deutsches Wörterbuch*, Neuausgabe, R. Wahrig-Burfeindt (ed), Gütersloh.
- WEIS Erich / MATTUTAT Heinrich, 1993. *PONS Großwörterbuch Französisch — Deutsch / Deutsch-Französisch*, Neubearb. unter Mitwirkung v. C. Nogue. Erw. Neubearb. 1. Aufl. v. 1988 (Nachdruck). Stuttgart/Dresden.
- ZIMMER Rudolf, 1990. *Äquivalenzen zwischen FRANZÖSISCH und DEUTSCH. Theorie — Korpus — Indizes. Ein Kontextwörterbuch*, Tübingen.